JamSketch: An Improvisation System with Melody Creation from User-drawn Melodic Outlines

Tetsuro Kitahara
College of Humanities and Sciences, Nihon University
Setagaya-ku, Tokyo, Japan
kitahara@chs.nihon-u.ac.jp

Sergio Giraldo
Music Technology Group, Universitat Pompeu Fabra
Barcelona, Spain
sergio.giraldo@upf.edu

Rafael Ramírez
Music Technology Group, Universitat Pompeu Fabra
Barcelon, Spain
rafael.ramirez@upf.edu

Abstract

We present a system called *JamSketch*, which enables the user to improvise collaboratively with the system by drawing melodic outlines on a touch screen. Improvisation requires advanced skill in creating melodies instantaneously, so many people cannot improvise even if they are skilled in playing an instrument. Our system enables such people to improvise based on the concept of "the macro structure by the user, the micro structure by the system." Once the user draws a melodic outline, which represents the overall shape (macro structure) of a melody, on a touch screen, the system immediately generates a melody based on the given melodic outline. The melody is generated using a genetic algorithm with a fitness function defined as a weighted sum of the similarity to the outline, N-gram probabilities (i.e., the similarity to a corpus of existing melodies), and an entropy. The N-gram is learned with a set of 53 Blues melodies. With this system, users can enjoy playing improvisational melodies with a given style (such as blues) even if they have no skills in creating melodies or playing an instrument.

Author Keywords

Improvisation; automatic melody creation; melodic outline; genetic algorithm

ACM Classification Keywords

H.5.5 [Information interfaces and presentation (e.g., HCI)]: Sound and Music Computing

Introduction

Improvisation is one of the most enjoyable forms of musical performance in which musicians create music compositions in real time by combining the communication of emotions, instrumental technique and spontaneous responses to other players. Hence, improvisation requires advanced skills in and knowledge of music. Many people cannot improvise even if they are skilled in playing an instrument. If computing technologies could enable these people to improvise without prior musical knowledge or skill, then they could enjoy a broader range of musical forms.

Here, we present a system that facilitates human-computer collaborative improvisation. The key concept of this system is "the macro structure given by the user, the micro structure given by the system." Once the user draws a melodic outline, which represents the overall shape of the melody that the user intends, the system creates a melody according to the outline in real time. Because drawing a melodic outline requires no musical skill or knowledge, people without musical skill or knowledge are able to enjoy improvisation.

System Overview

This system enables the user to input a melodic outline to play an improvisation. An example of melodic outlines is shown in Figure 1. While the backing track is played back, the user draws a melodic outline with his/her finger on a touchscreen built in or connected to the PC (Figure 2). Then, the system generates a melody that satisfies both the similarity to the melodic outline given by the user and musical typicality. Musical typicality is considered to avoid

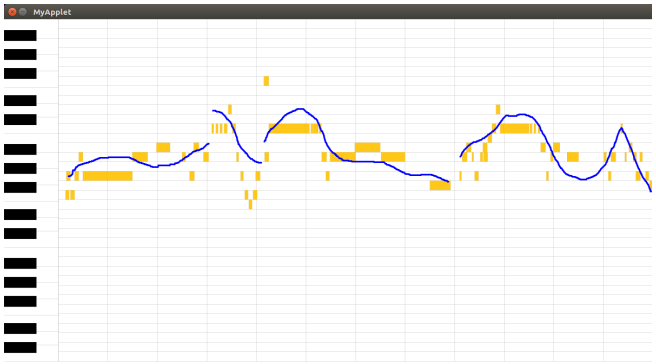

Figure 1: A example of melodic outlines (blue curves)

melodies that are musically inappropriate (e.g., cause dissonance).

Melodies are generated using a genetic algorithm (GA). Let $N = \{n_1, \cdots, n_L\}$ be a sequence of note numbers. The fitness function $F(N)$ for GA is defined as follows:

$$\begin{aligned} F(N) &= w_0\,\mathrm{sim}(N) + w_1\,\mathrm{seq}_1(N) + w_2\,\mathrm{seq}_2(N) \\ &\quad + w_3\,\mathrm{harm}(N) + w_4\,\mathrm{ent}(N), \end{aligned}$$

where

- $\mathrm{sim}(N)$: Similarity to outline

$$\mathrm{sim}(N) = -\sum_{i=0}^{L-1} \left(n_i - y(t_i)\right)^2,$$

in which t_i is the onset time of note n_i and $y(t_i)$ is a pitch at time t_i given from the melodic outline.

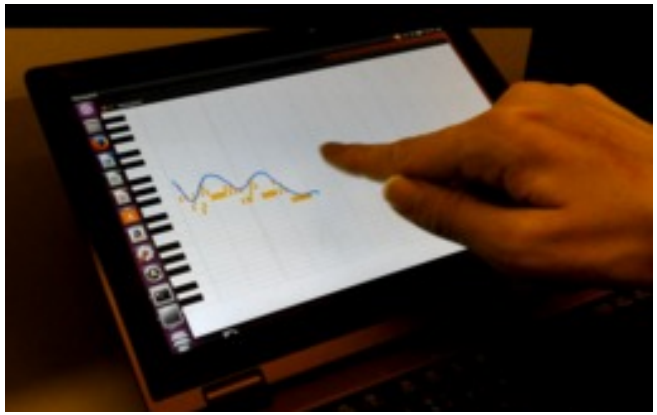

Figure 2: A scene of someone using our system

- $\mathrm{seq}_1(N)$: Pitch bigram probability

$$\mathrm{seq}_1(N) = \sum_{i=1}^{L-1} \log P(n_i \mid n_{i-1}).$$

- $\mathrm{seq}_2(N)$: Interval (pitch-motion) bigram probability

$$\mathrm{seq}_2(N) = \sum_{i=2}^{L-1} \log P(n_i - n_{i-1} \mid n_{i-1} - n_{i-2}).$$

- $\mathrm{harm}(N)$: Conditional probability for given chords

$$\mathrm{harm}(N) = \sum_{i=0}^{L-1} \log P(n_i \mid c_i, b_i),$$

in which c_i is the chord name at time t_i, and b_i is the metrical position at t_i ($b_i \in \{\text{head}, \text{on-beat}, \text{off-beat}\}$). We consider b_i because the acceptability of out-of-scale notes depends on their metrical positions.

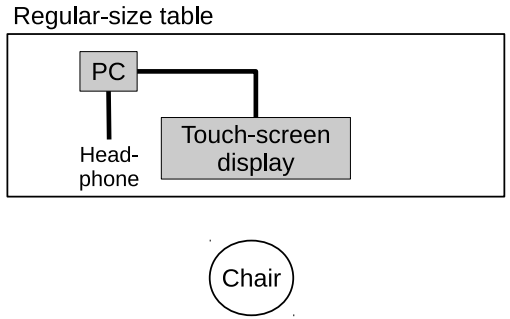

Figure 3: Technical specification

- $\mathrm{ent}(N)$: Entropy

$$\mathrm{ent}(N) = -(H(N) - H_{\mathrm{mean}} - \varepsilon)^2,$$

in which $H(N)$ is the entropy of $\{n_0, \cdots, n_{L-1}\}$, and H_{mean} is the averaged entropy calculated from a melody corpus. ε is usually zero, but setting this to greater than zero will result in more complex melodies.

In the above parameters, conditional probabilities including N-grams, that is, $P(n_i \mid n_{i-1})$, $P(n_i - n_{i-1} \mid n_{i-1} - n_{i-2})$, and $P(n_i \mid c_i, b_i)$, are learned from a corpus to imitate the typical style of a particular melody set. Here, we use 53 melodies with the tonality of Blues taken from the Weimar Jazz Database[1].

The user can select only a typical 12-bar chord progression: | C7 F7 C7 C7 F7 F7 C7 C7 G7 F7 C7 G7 |. This can be changed by preparing a MIDI file and a chord transcription.

Technical Specification

The system consists of a personal computer (PC), a touch-screen display, and a headphone (or a speaker) (Figure 3). Once a participant draws melodic outlines on the touch-screen display, the program running on the PC generates melodies and sounds them through the headphone (or the speaker). The system can be set up enough within the space of a typical table and does not require any special conditions about the power supply, lighting, or auditory environment. It takes less than 30 minutes to set up the system.

How we will present this work

We are planning to show this system to the participants as an interactive demo, and it may also be presented as an installation in order to enable the participants to improvise with this system for as long as possible.

Link to online video

A demo video is available at:
http://www.kthrlab.jp/proj/jamsketch/videos/demo.mp4

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 16K16180, 16H01744, 16KT0136, and 17H00749, as well as by Kawai Foundation for Sound Technology and Music, the Spanish TIMuL Project (TIN2013-48152-C2-2-R), and the TELMI Project of the Horizon 2020 Research and Innovation Programme (grant agreement No. 688269).

[1] http://jazzomat.hfm-weimar.de/dbformat/dboverview.html